R E P O R T R E S U M E S

ED 003 555 56

AN INVESTIGATION OF MOTION PICTURE FILM AND THE PROGRAM ANALYZER FEEDBACK TO IMPROVE TELEVISION TEACHER TRAINING.

BY- JOHNSON, F. CRAIG AND OTHERS
OHIO UNIV., ATHENS

REPORT NUMBER NDEA-VIIA-374 PUB DATE APR 60
EDRS PRICE MF-$0.09 HC-$1.48 37P.

DESCRIPTORS- *AUDIOVISUAL INSTRUCTION, INSTRUCTIONAL INNOVATION, *TEACHER EDUCATION, PROGRAM EVALUATION, *TELEVISION TEACHERS, *SOUND FILMS, AUDIOVISUAL PROGRAMS, INSERVICE TEACHER EDUCATION, FACTOR ANALYSIS, KINESCOPE RECORDINGS, ATHENS, OHIO

THE PURPOSE OF THIS STUDY WAS TO EVALUATE THE PROGRAM ANALYZER, KINESCOPE, AND AUDIOTAPE AS AIDS IN TRAINING TELEVISION TEACHERS. FOLLOWING 10-MINUTE PRESENTATIONS, EACH TEACHER'S PERFORMANCE WAS REVIEWED BY ONE OF FIVE METHODS--(1) SOUND ONLY, (2) SOUND AND PROGRAM ANALYZER, (3) KINESCOPE ONLY, (4) KINESCOPE AND PROGRAM ANALYZER, OR (5) SELF-EVALUATION ONLY. EACH LESSON WAS PRESENTED THREE TIMES WITH FEEDBACK FROM THE REVIEW FOLLOWING THE FIRST AND SECOND PRESENTATIONS. A SECOND LESSON WAS PREPARED AND PRESENTED FOLLOWING THE PRESENTATIONS OF LESSON ONE. THE RATING GROUPS WERE DIVIDED SO EACH MEMBER OF GROUP A REVIEWED ALL THREE PERFORMANCES OF GROUP B PRESENTATIONS, AND MEMBERS OF GROUP B REVIEWED ONLY A SINGLE PERFORMANCE OF GROUP A PRESENTATIONS. FINDINGS SHOWED THERE WAS NO DIFFERENCE IN RATINGS OF TEACHERS WHO (1) USED KINESCOPE, (2) USED AUDIO TAPE ONLY, OR (3) HAD NO REVIEW. RATINGS ON THE SECOND PRESENTATIONS OF TEACHERS USING THE PROGRAM ANALYZER WERE SIGNIFICANTLY LOWER THAN RATINGS OF THOSE WHO DID NOT USE THE ANALYZER. THE RATING MEANS FOR BOTH OF THESE GROUPS WERE EQUAL ON THE THIRD PRESENTATION, AND THE ANALYZER GROUP WAS NOT SIGNIFICANTLY BETTER ON THE FOURTH PRESENTATION. NO SIGNIFICANT INTERACTION WAS SHOWN TO RESULT FROM THE DIFFERENCES OF NUMBERS OF PRESENTATIONS REVIEWED BY THE TWO RATING GROUPS. (AL)

ED003555

U. S. DEPARTMENT OF HEALTH, EDUCATION AND WELFARE
Office of Education
This document has been reproduced exactly as received from the person or organization originating it. Points of view or opinions stated do not necessarily represent official Office of Education position or policy.

374

AN INVESTIGATION OF MOTION PICTURE FILM AND THE PROGRAM ANALYZER FEEDBACK TO IMPROVE TELEVISION TEACHER TRAINING

F. CRAIG JOHNSON, Chief Investigator

ELIZABETH G. ANDERSCH

PAUL A. GAMES

EDWARD M. PENSON

Final Report

April, 1960

A RESEARCH PROJECT SUPPORTED BY TITLE VII
NATIONAL DEFENSE EDUCATION ACT, 1958
AND
OHIO UNIVERSITY, ATHENS, OHIO

AN INVESTIGATION OF MOTION PICTURE FILM AND THE PROGRAM ANALYZER FEEDBACK TO IMPROVE TELEVISION TEACHER TRAINING

F. CRAIG JOHNSON, Chief Investigator

ELIZABETH G. ANDERSCH

PAUL A. GAMES

EDWARD M. PENSON

Final Report

April, 1960

A RESEARCH PROJECT SUPPORTED BY TITLE VII

NATIONAL DEFENSE EDUCATION ACT, 1958

AND

OHIO UNIVERSITY, ATHENS, OHIO

ACKNOWLEDGEMENTS

In any research project of the scope of the one reported here, there are always many people who contribute their time and special talents. The recognition represented here does not express the full value of their contributions, but does acknowledge, at least, their most appreciated services. The following people did much to insure the completion of the study:

Allen Carter filmed the one hundred sixty presentations, played them back for the teachers' review and photographed the pictures for this report.

Caroline Creath handled all of the correspondence necessary to contact prospective participants.

Margaret Warne Epperson was the secretary for the study, kept all of the records, and did much in helping to keep morale high.

Arthur Hafer did all of the technical work connected with the experiment and kept the television laboratory operating with somewhat awesome proficiency.

Robert Harold kept the analyzer running during a hot and humid summer.

Carl Trivett took over the care of the analyzer in the fall and nursed it through the final experimental trials.

Special appreciation goes to the teachers who participated in the study for their conscientious efforts and their good-natured understanding of the problems connected with the study; to the many Speech I students who served as raters; and to the Speech I instructors who adjusted their own classroom schedules in order to adhere to the rigid rating schedule.

Finally the experimenters wish to give special recognition to the administration of Ohio University for its willingness to support this project in both spirit and substance.

TABLE OF CONTENTS

	Page
The Problem	5
The Purpose	5
Administrative Procedures	8
Selection of Teacher-subjects	9
Selection of Student-raters	11
Experimental Procedures	11
Assignment of Subjects to Groups	11
Training the Teacher-subjects	13
Training the Peer Raters	14
Instrumentation	15
Difficulties Encountered	16
Results	16
Peer Ratings	16
Student Mean PAR Ratings	18
Analysis of the Relationships between Peer and Student Ratings	20
Correlations between Peer and Student Ratings	23
Summary and Discussion	24
Appendices	27

TABLES

		Page
Table I	Professional Duties of Participants	9
Table II	Classification of Schools Represented by Participants	10
Table III	Distribution of Ages of Participants	10
Table IV	Distribution of Years of Teaching Experience	10
Table V	Intercorrelations of Mean Peer Ratings over the Forty First Trial Presentations, above the Diagonal; the Residual Correlations, below the Diagonal; and the Centroid Factor Loadings of the Variables	16
Table VI	Means of the Student Mean PAR Ratings of the Five Groups	18
Table VII	Analysis of Mean Student PAR Rating, Summary Table	19
Table VIII	Analysis of Variance of Mean Student PAR Rating across Trials 1-4	20
Table IX	Mean PAR Ratings by Peers and by Students for the A and B Groups of Teachers on Trials 1, 2, and 3	21
Table X	Analysis of Variance of Mean PAR Ratings by Peers and Students	21
Table XI	The Correlations between Peer and Student Ratings for the A and B Groups on the Three Trials	23

FIGURES

		Page
Figure 1	Mean PAR Peer and Student Ratings Averaged across the First Three Trials	22
Figure 2	Mean PAR Peer and Student Ratings Plotted Separately for the First Three Trials	23

PHOTOGRAPHS

	Page
Teacher Presentation	6
The Teacher Alone in the Studio	6
Sound-on-Film Recording Equipment	7
Peer Rating Group	7
Rating Box and Value Scale	8
The Program Analyzer	12
PAR Group Training Device	14
PAR and Sound-on-Film Review	17
Sound-on-Film Review	17
Second Review for PAR Sound-on-Film Group	24

THE PROBLEM

Pre-service training is of major importance at all levels of television instruction.[1] *The National Program in the Use of Television in the Public Schools* lists as one of its objectives the development of teacher talent.[2] The Chicago City Junior College Project concludes that, "The adaptation of the teaching situation to the opportunities and limitations of the medium requires preparation and creativity well beyond that afforded by normal allotments of time for classroom teaching."[3] New York University finds that teachers are quite unprepared to teach by television.[4] The military research workers are aware of this problem. Kanner states, ". . . a teacher who has been effective in the classroom does not always make a successful transition to television teaching."[5] Training the classroom teacher to make an effective transition to television is as vital as it is time consuming.

The program analyzer has been used to improve content in many forms of communication. Twyford reviews fifteen applications of the program analyzer to radio, television, films, plays and lectures. He concludes that his program analyzer yielded reliable data with a correlation of .64 with a test measure of amount of learning. He suggests that, "Rating summators are an audio-visual aid and should be used to evaluate instruction and to facilitate interaction and participation in the classroom."[6]

THE PURPOSE

The purpose of this study was to evaluate the program analyzer, sound-on-film, and audio-tape as aids in training television teachers. Sound-on-film recordings were made of forty classroom teachers who each prepared a television lesson and presented it three times over television. Then, a second lesson, which contained different content, was prepared and presented over television. These lessons were rated on a five-point teaching effectiveness scale by groups of the forty teachers (peer ratings) and by one hundred and sixty groups of approximately eight college freshmen (student ratings). The program analyzer was used to summate all teacher effectiveness ratings (PAR).

1. Allen, William H., "Research on New Educational Media: Summary and Problems." AUDIO-VISUAL COMMUNICATION REVIEW, Vol. 7, No. 2, (Spring, 1959), p. 95.

2. "The National Program in the Use of Television in the Public Schools," TEACHING BY TELEVISION, A Report from the Ford Foundation and the Fund for the Advancement of Education, May, 1959, p. 47.

3. THE CHICAGO CITY JUNIOR COLLEGE EXPERIMENT IN OFFERING COLLEGE COURSES FOR CREDIT VIA OPEN CIRCUIT TELEVISION, March, 1958, p. 39.

4. CLOSED CIRCUIT TELEVISION AS A MEDIUM OF INSTRUCTION, New York University, October, 1956.

5. Kanner, Joseph H., "Future Trends in Television Teaching and Research," AUDIO-VISUAL COMMUNICATION REVIEW, Vol. 5, No. 4, (Fall, 1957), p. 517.

6. Twyford, Loran C., "Profile Techniques for Program Analysis," AUDIO VISUAL COMMUNICATION REVIEW, Vol. 2, No. 4, (Fall, 1954), p. 261.

TEACHER PRESENTATION. The fixed position videocon camera picks up this teacher as she makes a ten-minute presentation.

THE TEACHER ALONE IN THE STUDIO. Each teacher was responsible for the timing of his own presentation. A red light on the side of the videcon camera signified the start. The stop-clock in the center gave the performer the total running time.

SOUND-ON-FILM RECORDING EQUIPMENT. This monitor and camera sound-on-film combination served to record all one hundred sixty ten-minute presentations.

PEER RATING GROUP. In this air-conditioned room, groups of ten teachers rated the ten-minute television presentations of their fellow students.

The experiment included four conditions of feedback between presentations as shown on the diagram below, plus a control group that received no review of their presentation (no feedback).

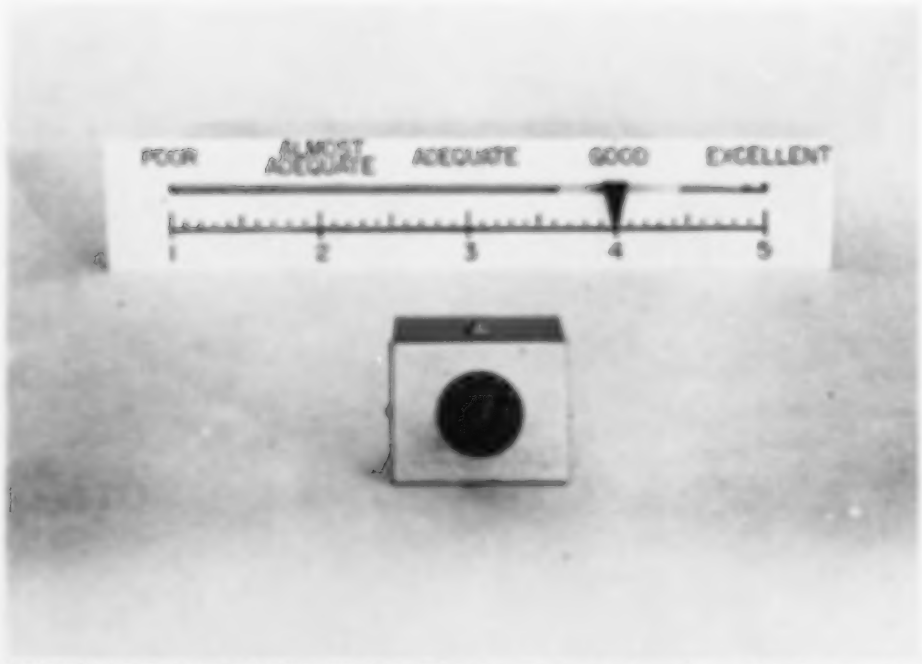
	No PAR	PAR
Film and Audio		
Audio only		

The experiment was designed to discover what effect if any, these varying feedback conditions would have on four successive television performances as judged by student ratings of teaching effectiveness. In addition, the effect of differential rating conditions on rating behavior was studied for two peer groups.

ADMINISTRATIVE PROCEDURES

Administrative implementation of the experimental design required decisions to be made in several broad areas. The decisions will be discussed under the following headings:

1. Selection of Teacher-subjects
2. Selection of Student-raters



RATING BOX AND VALUE SCALE. This detail shows the relationship between the five-position switches used by the raters and the value scale used to report the mean to the PAR GROUPS.

SELECTION OF TEACHER-SUBJECTS

A brochure announcing an Educational Television Teacher Training Institute and inviting application (Appendix A) and a cover letter (Appendix B) were prepared and distributed to approximately 1400 secondary school principals and 140 superintendents in the state of Ohio. In addition to this circularization by mail, members of the project staff personally contacted administrators in key cities.

As a result of the various efforts made to recruit participants, approximately 100 inquiries were processed. Each individual who responded to the initial contact was sent a copy of the brochure, an application blank (Appendix C), a list of courses which he was eligible to take, and a brief personal cover letter.

The Educational Television Teacher Training Institute extended over a five-week period scheduled to coincide with the second summer term at Ohio University. Three semester hours of graduate or undergraduate credit were allowed if desired. Students were permitted to elect an additional three-semester hours of course work.

Sixty-two invitations (Appendix D) to participate in the Institute were issued prior to June 15. As of June 27, forty-two persons had accepted. Prior to the opening of the Institute, forty-four persons had committed themselves to participation. However, by registration day, five teachers had dropped out of the program. Three individuals, on campus for the regular summer session, became interested in the program, applied for admission and were added to the group. Thus, the Institute began with forty-two participants: forty of these were considered subjects in the experiment and two were classified as alternates.

Of the forty teachers who served as subjects in the experiment, thirty-eight were associated with public schools in the state of Ohio; two participants were from out-of-state, one coming from Indiana and another from New York.

The majority of the participants taught in senior high schools, while the rest were divided among junior high schools and elementary schools, with one person operating as a supervisor in a city school system.

TABLE I

Professional Duties of Participants*

Type of School	Number	Percentage
Junior High School Teacher	2	5.0
Senior High School Teacher	33	82.5
Elementary Teacher	4	10.0
City Supervisor	1	2.5
Totals	40	100.0

In addition to their regular teaching assignments, several participants carried other responsibilities. One teacher also served as principal of his school; one was an art supervisor; one had supervisory responsibilities at the elementary level. Also, one of the persons teaching at the junior high school level had a position which required that she teach all subjects.

Various types of schools were represented by the forty participants, as indicated in the following table.

*The analysis of the data pertaining to the participants is based on pertinent information from the 40 subjects whose performances constituted the basic data of the experiment.

TABLE II

Classification of Schools Represented by Participants

Type of School	Number	Percentage
City High School	17	42.5
County High School	12	30.0
Exempted Village High School	3	7.5
City Elementary	3	7.5
City System-Supervisory Position	1	2.5
Military Academy	1	2.5
City Junior High School	1	2.5
County Junior High School	1	2.5
Township Elementary School	1	2.5
Totals	40	100.0

There were 14 male participants (35%) and 26 female participants (65%). The average age of the group was 43.1 years.

TABLE III

Distribution of Ages of Participants

Age	Number	Percentage
21 to 25	5	12.5
26 to 30	2	5.0
31 to 35	3	7.5
36 to 40	2	5.0
31 to 45	6	15.0
46 to 50	9	22.5
51 to 55	10	25.0
56 to 60	3	7.5
Totals	40	100.0

With the exception of one participant, all of the subjects had earned bachelor's degrees. Eighteen of the participants also possessed master's degrees. One individual had earned two bachelor's degrees.

The average number of years of teaching experience was 15.11 years, with a range of from one to forty years.

TABLE IV

Distribution of Years of Teaching Experience

Years of Experience	Number	Percentage
1 to 5	9	22.5
6 to 10	10	25.0
11 to 15	5	12.5
16 to 20	3	7.5
21 to 25	3	7.5
26 to 30	6	15.0
31 to 35	3	7.5
36 to 40	1	2.5
Totals	40	100.0

During the academic year of 1959-60, the forty participants taught a wide variety of subjects. In order to achieve the desired variety of television content some teachers were asked to select content for which they were academically prepared but were not presently teaching. Twenty-eight participants or 70% of the group taught television lessons in fields which they were currently teaching. Twelve participants or 30% of the total experimental group taught lessons in fields which they were not currently teaching but in which they were academically prepared, and in some cases, subjects which they had had experience in teaching or expected to teach.

Since each participant was to serve as a rater, and in view of the average age of the participants (43.1 years), the hearing ability of the subjects was determined by administering audiometric screening and threshold tests. Fourteen subjects were found to have significant hearing losses. This is not an unusually large number considering the ages of the subjects. Seven of these subjects had essentially normal hearing in one ear and this monaural handicap could not have been sufficient to interfere with either their rating or performing-with-feedback efforts. The remaining seven subjects did have severe biaural mixed or perceptive type hearing losses. These subjects were almost evenly distributed among the various groups involved in the project.

SELECTION OF STUDENT-RATERS

During the first semester, there were approximately 1250 freshmen enrolled in a basic speech course designated as Speech I in the Ohio University catalog. This course is required for graduation and usually is taken during the student's first or second semester. During the regular semester following the Institute, there were 87 sections of Speech I.

Because of the course organization, it was possible to arrange for these students to serve as evaluators without seriously interrupting the activities of the course. The students were scheduled to view the films in groups of seven or eight; in other words, one-half of a class, not committed to a performance assignment on a given day, viewed one of the films made during the summer. In all, one hundred and sixty groups were randomly assigned to view one hundred and sixty films.

EXPERIMENTAL PROCEDURES

ASSIGNMENT OF SUBJECTS TO GROUPS

Each of the forty teachers described above was asked to prepare a ten-minute content-unit lesson which he could select from his own teaching experience or academic training. He was told the lesson was to be delivered over television to a high school senior audience.

Half of the teachers presented their lessons during the first two weeks (Group A) and the other half during the second two weeks (Group B) of the summer session. While Group A was presenting their lessons, Group B served as raters and, conversely, while Group B presented their lessons, Group A rated.

When Group B rated Group A's performances, they were divided into three groups. Each group rated only one performance. When Group A

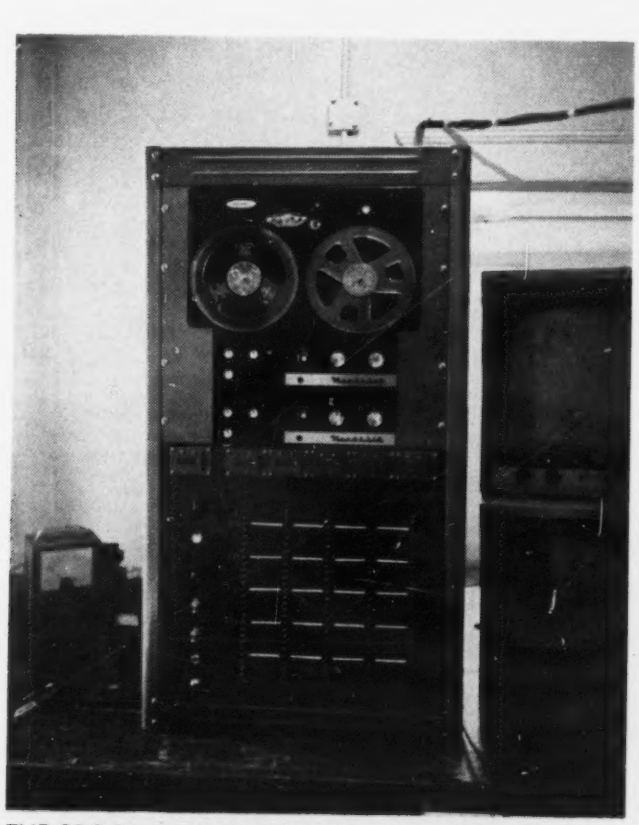
rated Group B's performances they were divided into two groups. Each group rated all three performances of one-half of the B teachers. The question was, would the rating behavior of these groups differ systematically as a result of the differing experiences.

When rating, then, Group A was divided into two rating groups (A_1 and A_2) and Group B into three rating groups (B_1, B_2, and B_3).

Each teacher reviewed his presentation under one of the following five conditions:

1. Sound only
2. Sound and PAR
3. Sound-on-film-only
4. Sound-on-film and PAR
5. Self-evaluation with no outside stimulus

Eight teachers (four from Group A and four from Group B) were randomly assigned to each of the five conditions.



THE PROGRAM ANALYZER. This close-up shows the basic rating-reporting instrument. The five colums of lights are used to report the number of boxes of a given position for a specific time segment. The tape recorders shown at the top were not available nor necessary for this study.

Each teacher-subject, then, was asked to serve two functions. He was to present a lesson over television and rate the presentations of his fellow teacher-subjects.

In the fifth week, after all forty teachers had presented their lessons three times and had received the feedback condition of their assigned group, they were asked to prepare a second lesson. This lesson was also ten minutes long, but contained different content again selected from their own teaching experience or academic training. This allowed comparisons to be made which reflected the transfer of the various feedback conditions.

TRAINING THE TEACHER-SUBJECTS

All teachers were given a brief orientation lecture on television presentation. They were told that one stationary camera would include a waist shot of them and the entire blackboard area behind them. They were to perform as they would in a familiar classroom situation, talking loud enough so students in the back row could hear, and using visual aids which would be appropriate for regular classroom use.

Each teacher was responsible for timing his own presentation. Instruction was given regarding the operation of the electric stop clock. Each teacher was given an opportunity to become accustomed to the performance conditions prior to the actual presentation.

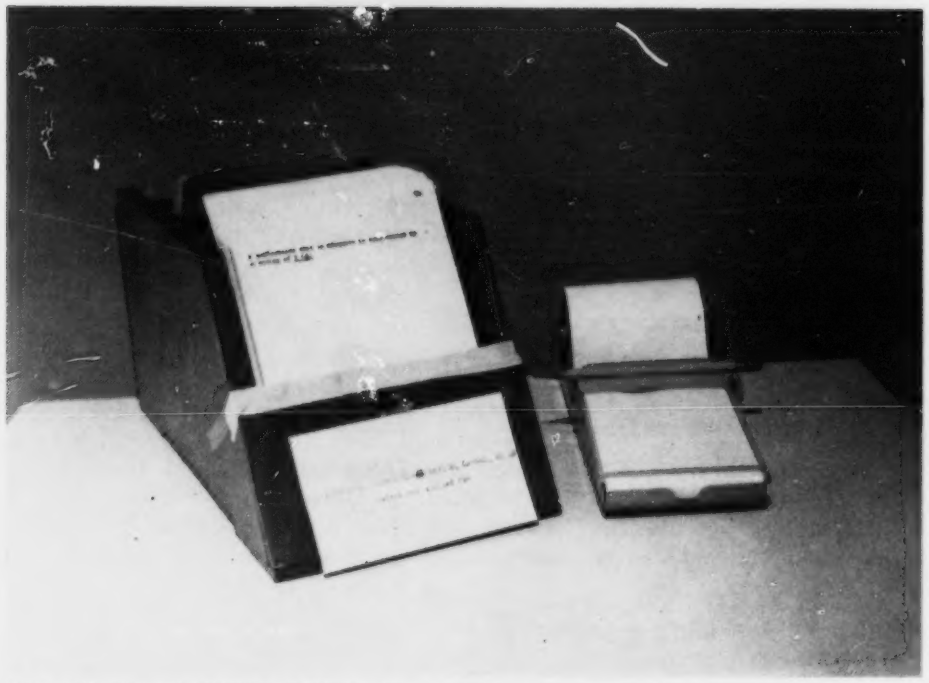
Special instructions were given to those teacher-subjects who were to use the PAR. Since the mean rating was not available for this study, the members using the program analyzer needed to learn to estimate a mean from the distribution of ratings the analyzer presented. To accomplish this, a simplified teaching machine was developed. Approximately seventy-two questions were prepared and typed on cards. The answers were covered and exposed after the subject had given his own answer.

No special instructions were given to the members of the other groups. All subjects were told they would be exposed to their aspect of feedback twice before they could make the subsequent presentations. They were asked to use whatever information they were able to acquire to improve their next presentations.

Between the presentations each group received the appropriate feedback twice. For the second exposure of the PAR groups, the mean rating was given using a simulated mean reporting device. This device consisted of the five-point scale with labels corresponding to the descriptive titles suggested to the raters:

1. Poor
2. Almost adequate
3. Average
4. Good
5. Excellent

Ten segments between each of the ratings facilitated the observation of decimal point readings. A pointer, operated by remote control by the supervising staff member, was moved to the appropriate reading (mean) for each time segment.



PAR GROUP TRAINING DEVICE. Those teachers in the PAR groups needed to make a quick estimate of their mean scores from the visual distribution of ratings shown on the Program Analyzer. The device pictured above was used for this purpose. It consisted of seventy-two questions and hidden answers in a programmed sequence.

TRAINING PEER RATERS

A filmed half-hour rater training lecture[7] was prepared. The film was shown to each rating group. The main points developed in this film were:

1. Use your own judgment as to the teaching effectiveness of the presentation.
2. The rating is to aid the teacher.
3. You will rate forty-five second segments while the red light next to the television set is lighted. Base your judgment on the previous forty-five second segment.
4. Try to have your rating reflect your true feelings by avoiding the tendency to:
   a. Rate all segments low.
   b. Rate all segments high.
   c. Give all segments an average rating because you can't decide.
   d. Rate all segments the same.

During the training period, two practice films[7] were presented. These films were specially prepared to include some examples of superior, average, and inferior teaching.

7. Films are on file at the New Educational Media Branch, Office of Education, Department of Health, Education, and Welfare, Washington 25, D. C.

After each performance was rated, the peer raters were asked to fill out a general evaluation sheet. Judgments were to be made on the overall teaching effectiveness, delivery, content and interest. The overall ratings were made by placing a mark on a graphic scale with the same verbal descriptions characterizing the five reference points on the scale (Appendix E). The marks were scored on a twenty-one point scale by subdividing the line.

INSTRUMENTATION

1. PROGRAM ANALYZER

Thirty, five-position switches provided the viewing audience with the five point scale described previously. Each position represented a different amount of electrical current. The current activated two kinds of reporting instruments. The first was a series of counters which were activated by stepper relays and reported the number of switches at a given position at a given moment. This number of switches was indicated by the number of lamps on the analyzer's panel which were lighted. These lights yielded the graphic distribution of the ratings. With this instrument, then, the speaker was able to hear and/or see his message and see the raters' responses. These responses were recorded and relayed to the analyzer's light panel every forty-five seconds. The means were recorded manually on an individual "record sheet" (Appendix F) and later presented to the individuals in the PAR group.

2. FILM

The film used for this study was the Eastman Kodak Plus X, B wind, reversal film (number 7276). An ASA exposure index was used to determine all exposures. Contrast was adjusted on the television monitor for a one to eight brightness range. The reversal processing of the film was done according to the recommendations of the Eastman Kodak Company. The actual processing of the film was accomplished by Valdhere Films Inc., of Dayton, Ohio.

An Auricon model NR-40 recording amplifier was used to produce variable area sound on the film. Input number three (600 ohm impedance) was used with the tone control set for speech only. Sound track exposures were set at "15;" the sound track noise reduction was set at "20." A seven-volt lamp was used in the camera.

3. TAPE RECORDER

The subjects (eight) who received only an audio playback of their message presentations listened to a Revere tape recorder using a standard magnetic tape (Scotch 111A).

4. STOP CLOCK-TIMER

In order to help the subjects terminate their presentations on time, that is, at the end of ten minutes (400 feet of film), a twelve-inch wall model stop clock was placed on a stand directly in front of each talker, but out of camera range. This timing device had a sweep hand which operated at one revolution per minute. In an effort to remind the subjects that they were to speak ten minutes, a special cardboard cut-out was super-imposed on the face of the clock so that only the first ten minutes of the sixty-minute face was clearly visible to the subjects.

DIFFICULTIES ENCOUNTERED

The difficulties which arose during the conduct of the experiment are discussed below.

1. The sound strip was lost on one 1200 foot segment of film involving the third presentations of three subjects in Presentation Group B. This exigency was adjusted by the process of "lip-syncing" the audio tape recording with the film. The raters reported no distraction during the presentation of these "lip-synced" segments.
2. The first five minutes of the second presentation of one subject in Presentation Group A had a defective sound strip. The same procedure as described under "1" above was employed to rectify the situation. Again, there were no comments forthcoming from the raters which would suggest they were distracted or even aware of a difference in lip and sound synchronization.

RESULTS

PEER RATINGS

Several measures were obtained on each performance from the peer rating group. The mean PAR rating, the mean content rating, the mean delivery rating, the mean interest rating and the mean teaching effectiveness rating for each performance were correlated across the forty first trial performances. The following table of intercorrelations resulted. One factor was extracted leaving the corresponding residuals.

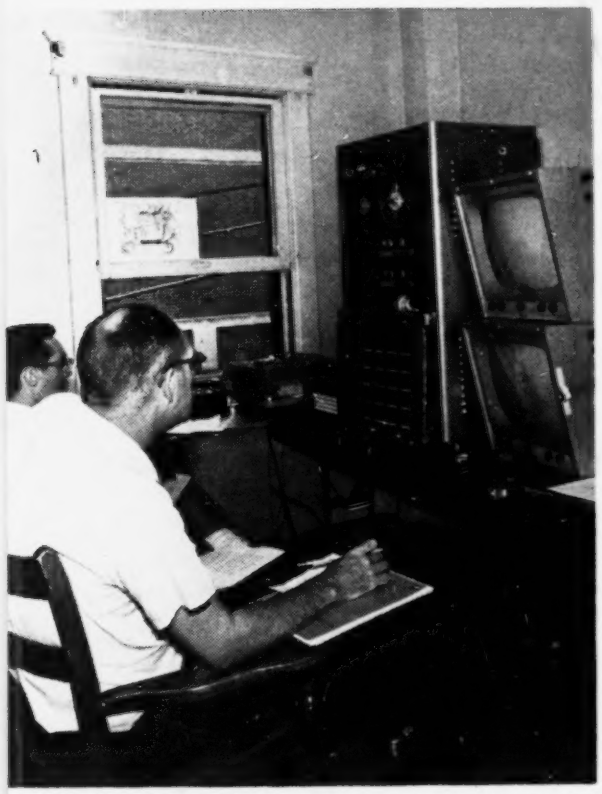
TABLE V

Intercorrelations of Mean Peer Ratings over the Forty First Trial Presentations, above the Diagonal; the Residual Correlations, below the Diagonal; and the Centroid Factor Loadings of the Variables

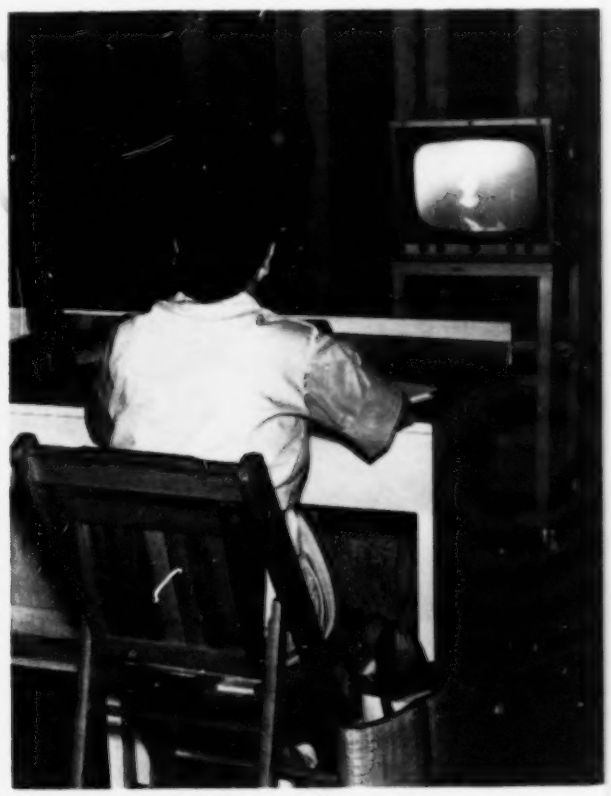
	PAR	TER	DR	CR	IR	Centroid Factor Loading
Mean PAR		.90	.92	.86	.88	.95
Teach. Effect. Rating-TER	–.02		.88	.92	.91	.96
Delivery Rating-DR	.03	–.01		.83	.84	.93
Content Rating-CR	–.03	.03	–.04		.86	.93
Interest Rating-IR	–.01	.01	–.02	–.01		.93

These results strongly suggested that the various named ratings were only measuring one thing; one factor accounted for all or almost all of the common variance, hence little was to be gained in the extra time necessary to collect supplementary measures. Consequently no measures except the

PAR ratings were collected from student raters. The square of the factor loading of .95 (.90) also may be taken as a minimal estimate of the reliability of the mean PAR rating by peers. This reliability estimate is analogous to a parallel forms measure of reliability. Time pressures prevented obtaining a reliability measure of the student ratings; however, there is no apparent reason why (for a given group of raters) reliabilities for students should differ appreciably. However the experimental process of using different randomly assigned student groups to rate each performance introduced additional sources of error variance not adequately represented in the peer rating correlations above.



PAR AND SOUND-ON-FILM REVIEW. The teacher in the foreground is watching a sound-on-film recording of his presentation made two days earlier. The peer rating group is also watching this film in another room and their ratings for this segment are reported to this teacher on the lighted panel as: 2 ratings of "Almost adequate (2);" 5 ratings of "Adequate (3);" 2 ratings of "Good (4);" and 1 rating of "Excellent (5);" for an average rating of 3.2.



SOUND-ON-FILM REVIEW. This teacher is seeing a sound-on-film recording of her ten-minute television presentation. Teachers who received only the sound portion of their recordings reviewed their presentations under similar conditions but were unable to see their pictures.

STUDENT MEAN PAR RATINGS

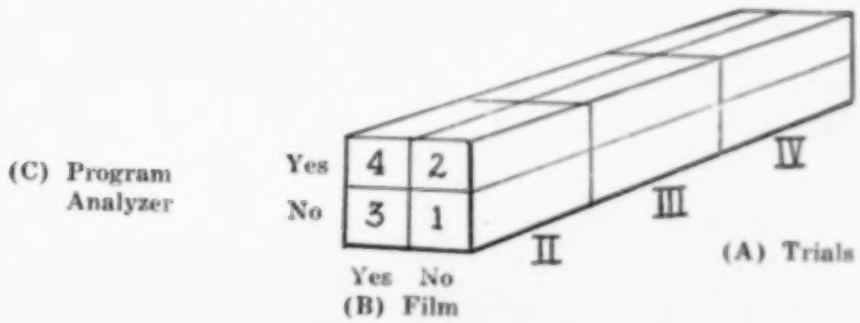
The mean student PAR rating was obtained on each of the four presentations of each teacher. The means of each group are presented in Table VI.

TABLE VI

Means of the Student Mean PAR Ratings of the Five Groups

Review Conditions	Trials			
	1	2	2	4
Audio only	2.81	2.95	2.91	2.88
Audio PAR	3.02	2.76	2.94	2.99
Film only	3.10	2.93	3.02	2.94
Film PAR	2.98	2.67	3.09	3.19
Control	2.89	2.80	2.90	2.95

The analysis was broken into two parts, a 2 x 2 factorial design repeated across trials (Lindquist, Type III[8]) and a two group design repeated over trials (Lindquist Type I[9]). The first analysis compared the following four groups across trials:



The first trial measures were obtained before any differential treatment of the groups occurred. This measure could be used as a control measure which would reflect the initial state of teaching effectiveness, providing the hypothesis of homogeneity of regression were retained. This hypothesis was retained ($F = .50$, $df = 3,24$), so the first trial measure was used as the control measure in an analysis of covariance of the between subject effects of means of trials 2, 3 and 4. The analysis of covariance of between subjects effects had the virtue of reducing the variability associated with initial teacher differences in teaching effectiveness, (the correlation within groups of the first trial measure with the subject sum over trials 2, 3 and 4 was .57). The within subjects effects were analyzed by usual analysis of variance techniques. The summary table of this analysis is presented below.

8. Lindquist, E. F., DESIGN AND ANALYSIS OF EXPERIMENTS IN PSYCHOLOGY AND EDUCATION, New York: Houghton Mifflin Co., 1956, p.
9. Ibid, p. 267.

TABLE VII

Analysis of Mean Student PAR Rating, Summary Table
Analysis of Covariance—between Subjects Effects

Source	df	adjusted ms	F
Audio-film (AF)	1	.002	<1
PAR	1	.016	<1
AF x PAR interaction	1	.351	<1
error (b)	27	.497	

Analysis of Variance—within Subjects Effects

Source	df	ms	F
Trials	2	.398	4.79*
Trials x AF interaction	2	.086	1.31
Trials x PAR interaction	2	.339	5.17**
Trials x PAR x AF interaction	2	.021	<1
error (w)	56	.066	

* Prob. <.05 ** Prob. <.01

The significant Trials x PAR interaction means that the trial effects differ for the two PAR groups; therefore, the ms trials were computed separately for the PAR groups and for the No PAR groups and tested against $ms_{s \times tr}$. The PAR groups trial effect was significant ($F = 9.65$, $df = 2,56$; Prob. <.001). The No PAR groups trial effect was not significant ($F = 2.21$, $df = 2,56$). Thus the only significant effect in the analysis was the change, across trials 2, 3 and 4 of the mean rating of the PAR groups. "t" tests of the equality of the pairs of means were run using the assumption of homogeneous trial-subject interaction effects. The mean of the second trial was significantly lower than the means of trials three and four (t's of 3.55 and 4.4 respectively with 56 df) but the hypothesis of the equality of the means of trials three and four was retained. Thus the results of the tests are that $M_{tr.2} < M_{tr.3} = M_{tr.4}$ for the PAR groups while $M_{tr.2} = M_{tr.3} = M_{tr.4}$ for the No PAR group. The audio-film dimension had no detectable effects; no terms involving this dimension were significant.

The second portion of the analysis compared the audio-only group with the control group. The same covariance-variance type analysis was proposed for these groups; however, the hypothesis of homogeneity of regression was rejected for these two groups ($F = 3.19$, $df = 1,12$; Prob. <.10) so a simple analysis of variance (Type I) was conducted. The analysis is summarized below.

TABLE VIII

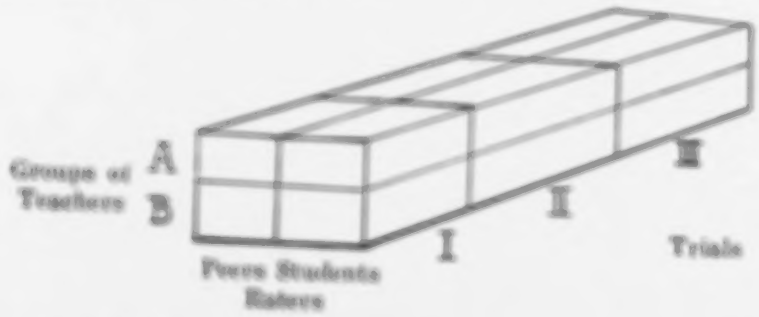
Analysis of Variance of Mean Student PAR Rating across Trials 1-4

Source	df	ms	F
Between Subjects			
Audio-Control	1	.0001	<1
Error (b)	34	.712	
Within Subjects			
Trials	3	.917	<1
Trials x AC interaction	3	.943	<1
Error (w)	42	.121	

(The corresponding analysis of covariance also yielded F's of less than 1.)

ANALYSIS OF THE RELATIONSHIPS BETWEEN PEER AND STUDENT RATINGS

Peer ratings were available for the first three presentations of each teacher, and student ratings were available for all four presentations. In an attempt to discover how the peer ratings differed from the student ratings and whether the differential experiences of the A and B groups of teachers affected their rating behavior, the following analysis of the first three trial ratings was conducted. The presence of Peer and Student ratings on every performance and the three trials by each teacher in the two groups, as diagrammed below, led to an analysis of variance by a design Lindquist refers to as Type VI.[11]



Recall that the 120 performances covered in this analysis were assigned in a random order to the groups of students (they were part of the total of 160 performances randomly assigned). There were no systematic administrative biases possible that could lead to any differential reactions of the students; therefore, their ratings may be conceived of as the "unbiased criterion ratings."

The A and B teachers, however, differed in sequence of rating and presentation (an administrative necessity to provide the necessary feedback) and the ratings they gave and the ratings their performances received might be affected by this difference in sequence. Recall that the A group gave all three performances prior to doing any rating; then each A

11. Ibid. p. 266.

rater rated all three performances of one-half of the B teachers. Eight of the twenty A subjects had seen the ratings of their own performances before they rated the other teachers. The members of the B group, on the other hand rated one performance of all members of the A group before they gave their presentations, and thus might have benefited by watching before they had to present. The means of the PAR ratings are given in Table X.

TABLE IX

Mean PAR Ratings by Peers and by Students for the A and B Groups of Teachers on Trials 1, 2, and 3

	Peer Ratings		Student Ratings		Overall
	A	B	A	B	
Trial 1	3.43	3.20	2.87	3.05	3.14
Trial 2	3.32	3.20	2.79	2.87	3.05
Trial 3	3.40	3.43	2.97	2.99	3.19
Overall	3.39	3.27	2.87	2.97	3.125

The results of the analysis are given below.

TABLE X

Analysis of Variance of Mean PAR Ratings by Peers and Students

Source	df	ms	F	Prob.
Between Subjects				
Groups (G)	1	.0032	<1	
error (b)	38	.9731		
Within Subjects				
Trials (T)	2	.4542	5.01	P<.01
Raters (R)	1	10.0123	79.85	P<.001
T x R int.	2	.0438	<1	
T x G int.	2	.0143	<1	
R x G int.	1	.6407	5.11	P<.025
R x T x G int.	2	.2252	2.66	.05<P<.10
$error_1$(w) [T and T x G]	76	.0907		
$error_2$(w) [R and R x G]	38	.1254		
$error_3$(w) [TxR and TxRxG]	76	.0848		

The main trials effect was the only term involving trials that was significant. The overall trial means were 3.14, 3.05, and 3.19 for trials 1, 2, and 3 respectively. The trial 2 mean was significantly lower than the trial 1 mean ($t = 2.75$, $df = 76$, Prob. $< .01$) and the trial 3 mean ($t = 4.43$, $df = 76$, Prob. $< .001$); however, the trial 1 and trial 3 means were not significantly different ($t = 1.68$, $df = 76$, Prob. $> .05$). This result is consistent with that found in the main analysis of student ratings where it was found that the trial 2 mean of the PAR groups was significantly lower that the other means. Since there were eight PAR subjects in the A

and in the B groups, the poorer performance given on the second trial lowered the mean of both groups. It is interesting to note that both the students and peers responded to this poorer performance by lowering their ratings, otherwise there would have been a raters by trials interaction.

Using the .05 level of significance, the only significant interaction is the main raters by groups interaction. This means that the A-B group effect differs from the Peer and the Student ratings. This interaction is plotted below.

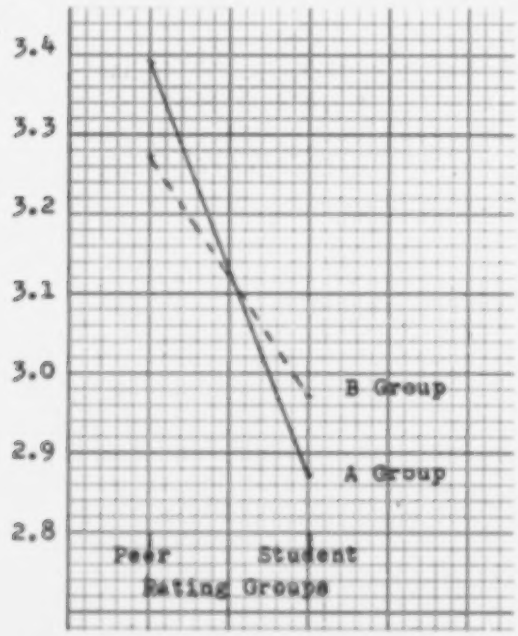


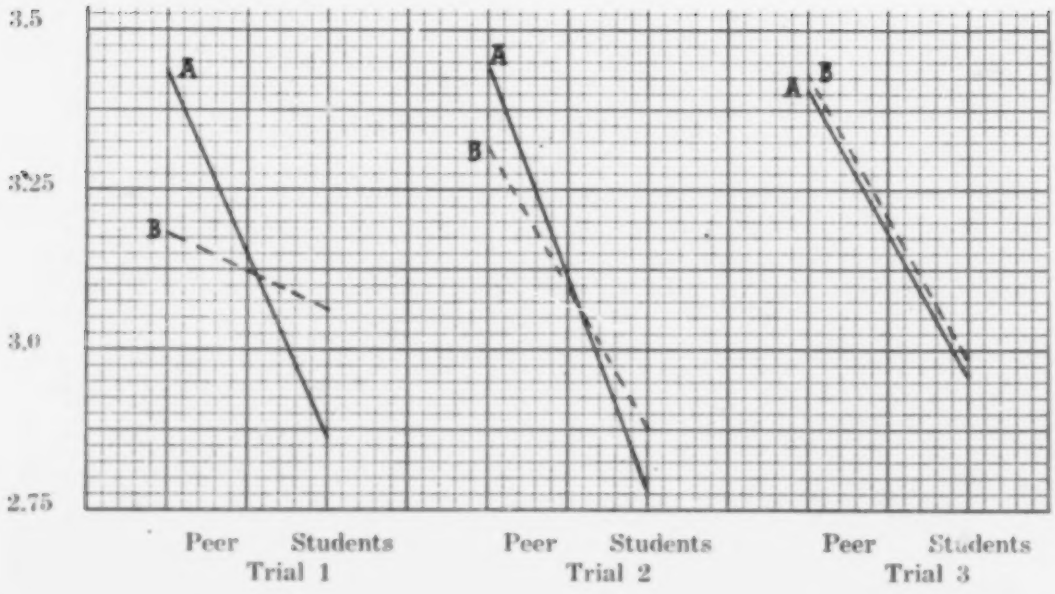
Figure 1. Mean PAR Peer and Student Ratings Averaged across the First Three Trials

We see that in the student ratings, the B group mean is higher than the mean of the A group, suggesting that this group may indeed have benefited from watching twenty A presentations before giving one presentation of their own. On the Peer ratings, however, the A group mean is higher than that of the B group.

Despite the fact that the interaction was significant, neither of the simple effect differences between the groups is significant; ie, the A and B groups are not significantly different for the Peer ratings alone ($F < 1$) or for the Student ratings alone ($F < 1$). Both of these tests, however, are "between subjects" tests and have a lower power than the test for interaction (a within subjects test). The significant interaction provides some justification for the conclusion that these means differ despite the lack of significance of the simple effects. The mean rating by Peers is greater than the mean rating by students for both the A group ($F = 62.67$, $df = 1,38$) and the B group ($F = 22.28$, $df = 1,38$).

The prob. of the obtained F on the triple interaction was greater than .05 and less than .10. If significant, this would have meant that the Rater x Group interactions differed from trial to trial. This effect was not significant. The obtained variation may be due to chance; however, since there is always the possibility of a type II error, perhaps a look at the trend of the data is worthwhile. The Group-rater means are plotted separately for the three trials.

Figure 2. Mean PAR Peer and Student Ratings Plotted Separately for the First Three Trials



Note that the Group-rater interaction was quite pronounced on trial 1, that it was reduced on trial 2, and nonexistant on trial 3. If separate F tests are run on the Group-rater interactions (assuming homogenious Rater subjects interactions and using $error_2(w)$ throughout) the F's are 7.02, 1.70 and .001 respectively with only the first trial interaction significant. (Similar results are obtained if the homogeneity assumption is not made and separate Type I analyses are run for each trial). With the lack of significance of the triple interaction, however, the only conclusion we are justified in making is that trking the subjects mean across the three trials, the B group scored lower than the A group on Peer ratings, but higher on Student ratings and this interaction can not be attributed to chance.

CORRELATIONS BETWEEN PEER AND STUDENT RATINGS

Since the Raters-Groups interaction indicated the means of the A and B groups were different for the two sets of raters, the Peers and the Students, it was decided to keep the A and B groups separate when computing the correlations between Peer ratings and student ratings.

TABLE XI

The Correlations between Peer and Student Ratings for the A and B Groups on the Three Trials

	Trial 1	Trial 2	Trial 3
Group A	.76	.49	.78
Group B	.65	.47	.49
Within Groups	.70	.48	.65

It is possible to test the equality of correlations on independent groups using Fisher's z transformation. Each A group correlation may be tested against the three B group correlations, making a total of nine such tests. None of these tests were significant; it should be noted, however, that these tests are not independent and, with only twenty subjects per correlation, the power of these tests is quite low.

SUMMARY AND DISCUSSION

1. Those teachers who reviewed their presentations using the program analyzer received significantly lower ratings on their second presentations than did those teachers who did not use the program analyzer. The means for both groups were equal on the third presentation and the program analyzer group did slightly (but not significantly) better on the fourth presentation. This trend produced a significant Trials-by-PAR interaction.

2. There was no significant difference between the ratings given the teachers who reviewed their presentations using a sound-film and those who used only an audio tape.

SECOND REVIEW FOR PAR SOUND-ON-FILM GROUP. The teacher on the right makes notes during his second review prior to the filming of his revised presentation. His mean ratings for each segment have been calculated and are reported to him by the manually operated sliding scale shown below his picture. This same room was used for the second review for the audio-only and PAR groups.

3. There was no significant difference between the ratings given the teachers who reviewed their presentations using an audio tape and those who had no review.

4. There was no other significant interactions.

DISCUSSION: STUDENT RATINGS

The major analysis of the four feedback groups and subsidiary analysis of the control group and audio group revealed only one significant effect—the significantly lower mean of the PAR groups on the second presentation. The mean rating went down considerably from the first trial, suggesting that the effect of observing peer ratings of the first performance actually disrupted the second performance. Exactly why this decrement in performance occurred is hard to assess. In any case, the decrement disappeared by the third presentation and by the fourth presentation, the PAR group ratings were above the remaining groups (however, this difference was not significant). Future studies should certainly contain more than four presentations to determine if desirable modifications of behavior of the PAR groups occur after "a period of adjustment" on the early trials.

Many of the teachers displayed some resistance against giving three presentations of the same material. Comments were made that it was like giving the same lecture three times. No objections were raised to the second presentation, but many apparently would have preferred to have presented new material on the third presentation. The reaction to the fourth presentation, when each did use new material, was far more favorable. Future studies should probably restrict themselves to only two presentations of each given content, or else, have a new content for each presentation.

5. There is a very high correlation between mean peer ratings (made during the presentation) of "teaching effectiveness" and mean peer ratings (made after the presentation) of overall "content," "delivery," "interest" and "teaching effectiveness."

DISCUSSION: PEER RATINGS

The high correlations between the so-called different ratings of "content," "delivery," "interest," "teaching effectiveness" and the PAR rating suggest that there is no point in asking members of this peer population to try and distinguish between the varying aspects of a presentation; they can't do it. The high loadings on the single centroid factor extracted, clearly imply that it does not matter what you call the dimension you ask them to rate; the teachers when rating apparently develop a general impression of the presentation and all ratings reflect this single general impression.

Part of the problem here may be that the subjects rated for a considerable period of time. Instructions were given at the very start of rating and these instructions attempted to stress that these ratings could differ considerably for a given performance. (i.e., delivery high, interest low, etc.) However, since the rating continued for a considerable period of time, the teachers may have eventually disregarded such instructions. Unfortunately, time prevented us from obtaining

similar data from student raters to determine if only one factor were present in the rating behavior of this population, also.

6. Mean student ratings of teaching effectiveness were significantly lower than the mean peer ratings.

DISCUSSION: COMPARISONS OF PEER AND STUDENT RATINGS

One comment applicable to both peer and student ratings on the PAR is that the means centered closely around three, the midpoint of the scale, and a rating of "average." This occurred despite very extensive instructions and a trial rating period designed to encourage the student to rate all the way from "five" to "one."

This is a common finding in studies where groups of people give ratings based on subjective criteria. In addition, performances typically varied enough in quality from time to time, so that the scores would tend to approach a middle value. It is very difficult to be "very poor" all of the time, and almost impossible to be "Excellent" all of the time. Since all statistical comparisons are based on the variability that actually existed in the data, this restriction of the scores does not handicap the study.

The most pronounced difference between the peer and student ratings is the greater leniency of the peers as raters, as displayed in the consistently higher means. The other interesting trends are in the Group-by-rater interaction. The only problem in interpreting this interaction is whether the interaction differs from trial to trial or not. The separate trial data looks as if the B group performed at a higher level on the first trial, as a consequence of having watched previous performances; but this superiority (as judged by student ratings) was reduced by trial 2 and almost non-existent by trial 3. On the other hand, on the Peer ratings, Group B started lower than Group A, moves closer on trial 2, and is very slightly above A on trial 3. This could possibly be a temporary reduction of the peer leniency tendency, maybe due to the experience of having just performed, or perhaps due to the effect of having been rated.

Over the three trials the B group was rated lower than the A by peers, while the B group was rated higher than the A group by students. These two effects combined to produce the significant Rater-by-groups interaction. The separate trial means suggest that this is a temporary effect, however; that is, maximal on the first trial, and lost by the third trial. Unfortunately the triple interaction, which should be significant to support the notion of this as a temporary effect, is in a relatively ambiguous position, with a Probability between .10 and .05. Thus, this change from trial to trial might be attributable to chance variation.

The correlations between peer and student ratings show some fluctuation from cell to cell of Table XI; however, none of the differences that can be tested were significant. Unfortunately, the small N and low power of this test leave much to be desired.

educational
television
teacher training
institute
with
ohio university
school of
dramatic art
and speech

APPENDIX A

• Purpose

A project to investigate certain techniques which may facilitate the classroom teacher's adaptation to the medium of television. This study will be conducted by the School of Dramatic Art and Speech in co-operation with the United States Office of Education and was made possible through a grant from its Department of Health, Education, and Welfare.

• Financial Arrangements

Participants selected will have all expenses for board and room paid. In addition, a stipend of $55 will be paid to each participant upon completion of the Institute.

• Tuition

$12 per credit hour or $55 for the summer session.

• Television Program

The Television Institute will be held during the second five-week summer session (July 18 to August 20) and students will register for Radio-Television 360—Instructional Methods in Educational Television—for three hours of credit.

• Credit

A participant may enroll for credit in an additional three-hour course selected from classes scheduled in the early morning hours. Thus, a total of six semester-hours may be earned during the five-week period.

Graduate status will be granted to any participant who secures admission, or has already been admitted to the Graduate College of Ohio University. One who holds a master's degree will be admitted automatically if he applies. One who currently has a graduate program in progress at another institution will be admitted automatically if his application is accompanied by a statement of good standing in this program. Otherwise, admission to the Graduate College is limited to those who have bachelor's degree with a cumulative 2.5 average (C—2.0), or better, from an accredited institution.

Appointment to the Institute does not guarantee admission to the Graduate College. Applicants for stipends are requested *not* to make application for admission to the Graduate College until they have been notified of appointment to the Institute.

Even though a participant receives graduate credit for work completed in the Institute, the application of this credit to a graduate program will usually require the approval of his graduate program advisers where the program is in progress.

• Housing and Meals

Participants in this program will reside in the same dormitory and will all eat at the same cafeteria. An offer of appointment to the Institute is a guarantee of a single room reservation.

• Recreation

Recreational facilities available to regular summer session students will be available on the same basis to Institute participants. These include golf, bowling, tennis, swimming, convocations, the Ohio Valley Summer Theater, free movies and concerts. Nearby state parks are accessible for outings.

• How to Apply

If you are presently engaged in teaching in senior high school, you are eligible for consideration for this Institute. If you are interested in making application, mail a postcard request at the earliest possible date to:

DR. E. G. ANDERSCH
Ohio University
Athens, Ohio

You will be sent the necessary application forms immediately.

APPENDIX B

COVER LETTER TO ADMINISTRATORS

OHIO UNIVERSITY

COLLEGE OF FINE ARTS

ATHENS, OHIO

SCHOOL OF DRAMATIC ART AND SPEECH
SCHOOL OF PAINTING AND ALLIED ARTS
SCHOOL OF MUSIC

February 15, 1960

To the superintendent or principal:

Many announcements of annual workshops, institutes, and seminars are sent out every year to various administrators of the public schools. We feel that the enclosed announcement of an Educational Television Teacher Training Institute is unique since it probably will be the only one of its kind in the nation.

The United States Office of Education and Ohio University are interested in the problem of training teachers to use the new medium of television. A research project costing some $57,600 has been designed to investigate some critical questions facing schools who are using or planning to use educational television.

For the most meaningful results, we need the cooperation of a group of secondary school teachers representing a variety of subject-matter-areas. We would appreciate your help in bringing this project to the attention of your faculty.

We feel that the Institute will be an interesting experience for the participants. Also, each individual will incur very little financial obligation since the Institute will pay the board and room for each and will pay each participant $55 at the conclusion of the 5-week period. Furthermore, each student may earn up to 6 hours of credit.

To a very large degree we must depend upon your help in securing individuals who would be interested in helping us with this project. Please ask applicants to write for an application blank to:

Dr. E. G. Andersch
Ohio University
Athens, Ohio

APPENDIX C

APPLICATION BLANK

EDUCATIONAL TELEVISION TEACHER TRAINING INSTITUTE

OHIO UNIVERSITY, ATHENS, OHIO

JULY 18 TO AUGUST 20, 1960

APPLICATION

(Please Type)

Name: _____

Present Home Address: _____

Name of School: _____

Check type of school system: County___, City___; Exempted Village___

In what county: _____

Address of School: _____

Present Position: _____
Subject
Grades

Name of superintendent: _____

Educational Background:

College or University Name of Institution	Bachelor's Degree Yes___ No___	Master's Degree Yes___ No___	Major	Minor

Experience:

Position	Place	Approx. Dates

Do you wish to register for: No credit___; 3 hours credit___; 6 hours credit___?

Do you wish to enroll for graduate credit? Yes____ No____

Signed_____

Position_____

Return to:
E. G. Andersch
Ohio University
Athens, Ohio

APPENDIX D

LETTER OF ACCEPTANCE

OHIO UNIVERSITY

COLLEGE OF FINE ARTS

ATHENS, OHIO

SCHOOL OF DRAMATIC ART AND SPEECH
SCHOOL OF PAINTING AND ALLIED ARTS
SCHOOL OF MUSIC

We are pleased to inform you that you have been awarded one of the grants for the Educational Television Teacher Training Institute which will be held by Ohio University during the second summer session from July 18th to August 20th. (Registration is on July 16th.)

Please indicate your decision in connection with this offer by checking the proper space on the second copy of this letter and returning it to us at your earliest convenience. Since we are accepting a maximum of only forty-four participants, should you decide that you cannot accept this offer, we need to have sufficient time to make this same offer to another applicant.

This offer is made possible through a grant from the United States Office of Education, Department of Health, Education, and Welfare and includes prepaid board and room and a stipend of $55.00.

Your decision should be based on a full understanding that the cost of tuition ($12.00 per hour or $55.00 for the summer session) must be paid by you. However, you will receive a stipend of $55.00 at the end of the term. The audit fee is the same as the fee charged for credit. Also, may we remind you that we are requiring all participants to live in a dormitory reserved for this project and to eat at a specified dining hall on campus.

Acceptance of this work for graduate credit must be cleared with the Graduate College. If you accept this appointment, your name will be sent to the Dean of the Graduate College who will send you the necessary forms for application for graduate credit.

If you accept this offer, please fill in the enclosed application for housing which should be returned to me for your room assignment. This offer of appointment to the Institute is a guarantee of a single room reservation.

Please feel free to make any announcement of this appointment to the press. Later in the spring we will send news releases to all Ohio newspapers.

We are looking forward to working with you this summer and trust that the program will be mutually helpful.

Sincerely,

Elizabeth G. Andersch, Professor of
Dramatic Art and Speech

To: Dr. E. G. Andersch, Ohio University, Athens, Ohio

_____I am pleased to accept the appointment to the Television Institute.

_____I respectfully decline the appointment to the Television Institute.

APPENDIX E

EVALUATION BLANK

Overall Evaluation Sheet—Performance_____

Subject #_____ Time _____

Put an X anywhere on the line to represent your overall estimate of each category. The X's may be inbetween numbers as well as on the numbers.

	poor	almost adequate	adequate	good	excellent
Content	1	2	3	4	5
Delivery	1	2	3	4	5
Interest	1	2	3	4	5

Teaching Effectiveness (How much did you learn?)

1_____2_____3_____4_____5

Did anything occur which prevented you from doing an accurate job of rating?

Yes No (Circle answer)

If so, what? _____

_____

_____

_____

Did the distraction occur during the *early, middle, or late* part of the presentation? (Circle which)
Any comments?

APPENDIX F

PAR RECORD SHEET

Name_____ Time of presentation_____

Review Times _____ _____ Presentation #_____

Period	Ratings									Period
	1	2	3	4	5	ΣX	ΣX^2	σ^2	X	
1										1
2										2
3										3
4										4
5										5
6										6
7										7
8										8
9										9
10										10
11										11
12										12
13										13
14										14
15										15
16										16
17										17
18										18
19										19
20										20
Σf										

Notes: